International Journal of Computational Methods and Experimental Measurements
Vol. 12, No. 3, September, 2024, pp. 269-279
Journal homepage: http://iieta.org/journals/ijcmem

Indoor Sound Classification with Support Vector Machines: State of the Art and Experimentation

Leila Abdoune[1,2*], Mohamed Fezari[3], Ahmed Dib[4]

[1] Computer Science Department, University Badji Mokhtar Annaba, Annaba 23000, Algeria
[2] Computer Science Department, ENSET (Ecole Normale Supérieure de l'Enseignement Technologique de Skikda), Skikda 21000, Algeria
[3] Electronics Department, Faculty of Engineering, University Badji Mokhtar Annaba, Annaba 23000, Algeria
[4] Networks and Systems Laboratory - LRS, Department of Computer Science, University Badji Mokhtar Annaba, Annaba 23000, Algeria

Corresponding Author Email: lilouchetoday@yahoo.fr

https://doi.org/10.18280/ijcmem.120307

Received: 8 August 2024
Revised: 16 September 2024
Accepted: 25 September 2024
Available online: 30 September 2024

Keywords:
sound recognition, sound classification, support vector machines, indoor sounds, Mel frequency cepstral coefficients, abnormal sounds, surveillance system

ABSTRACT

Sound classification is considered as one of the most important areas of classification domain, but the least developed compared to speech and voice recognition. In this study, we focus on the works that deal with sound classification by making a comparative study based on feature extraction and classification methods as well as the targeted sound corpus. Next, we present an overview of sound classification systems that utilize deep learning techniques, aiming to compare them with traditional learning methods. Based on our previous studies and conclusions, and considering that the challenge in choosing classification methods lies in balancing accuracy and computational cost, we conducted experiments using SVMs (support vector machines) with different kernels and MFCCs (Mel frequency Cepstral coefficients). Tests are carried out for the classification of some indoor abnormal sounds, then the number of classes is increased to cover a wider variety of sounds in order to observe and study the system's behavior. Finally, the results obtained in this work are promising and motivate us to explore deeper tests which are mentioned in the discussion and conclusion section.

1. INTRODUCTION

In daily life and anywhere we go, we encounter a variety of sounds such as birdsong, shouts, the clinking of dishes, phone ringtones, and footsteps. This collection of sounds varies from one place to another and even varies within the same place depending on the scenario or activities taking place. The ability to automatically detect these sound events is called Automatic Sound Recognition (ASR). Other terms can also be used to indicate sound recognition, such as Acoustic Event Recognition (AER) and Audio Event Recognition. Recognizing environmental sounds or, specifically, recognizing everyday life sounds is a particular case within the field of sound recognition.

The majority of research has focused on speech and speaker recognition; however, work on non-speech environmental sound recognition has been limited. Nevertheless, these efforts have become more active in recent decades [1]. More recently, there has been a growing interest in the classification of acoustic scenes [2-4].

An important characteristic of non-speech sounds is their diversity and the variation in their structures. Therefore, existing recognition systems are specialized in dealing with a particular type of sound or a set of sounds but are unable to handle all categories of sounds. This has made it challenging to find the right methods for Sound Event Recognition. As a result, this field remains open for finding solutions, comparing and attempting to adapt existing methods, and even discovering new methods.

In fact, in a real environment, the sounds to be recognized may be noisy, i.e., with the presence of environmental noise, or overlapping, indicating several sounds that are nested and superimposed. Consequently, to classify them, we first need to remove the noise by extracting the sound from the background noise [5], and then separate the sounds so that we can process them separately. All this processing is an extension of sound recognition and forms what we call Auditory Scene analysis (ASA) [6].

Several fields can benefit from sound recognition such as robotics, hearing aid technologies, security systems, remote monitoring and assistance systems and also disaster response, to name a few. Although there is a significant amount of work in this area that provides good recognition rates, there are still several major challenges to be addressed including the diversity of sounds, their overlapping nature, the very varied structure of sounds that vary from stationary, quasi-stationary to non-stationary, in addition to impulsive, each requiring specific methods, in addition to the lack of a standard database due to the large amount of sounds, etc.

Indeed, researchers in the field of sound event recognition other than speech and music are divided into two groups [7], one group that tries to exploit, borrow and adapt the feature extraction methods used in the fields of speech and music recognition which are frequently stationary techniques such as MFCCs that is originally used for speech and music recognition. However, the second group try to use specific feature extraction methods that take into account the properties of the environmental sounds and develop domain-specific techniques to capture unique sound characteristics, like non-stationary techniques. While the non-stationary methods give improved performance, they are often computationally expensive. On the other hand, although stationary techniques are easy to compute but the modeling of non-stationary sounds faces certain limitations [8]. So, we have to be careful in the choice of methods to achieve a balance between performance and computation time.

This work focuses on the recognition of environmental sounds as isolated sounds produced within an apartment (indoor sounds), with the aim of integrating them into a more comprehensive real-time sound recognition system that serves as a surveillance system. In general, like any problem of recognition and machine learning, three important points must be addressed: the choice of feature extraction methods, classification methods, and finally, the database for benchmarking and testing the system. These are critical points for making decisions on the performance and accuracy of the system to be developed. Indeed, the real challenge in sound recognition lies in the selection of feature extraction methods, as well as finding the most suitable methods for recognizing these sounds taking into account the nature of the intended application (real-time processing, required processing power, potential integration into the Internet of Things (IoT)).

The aim of this work is primarily to show the possibility of using feature extraction methods used for speech and music recognition, such as MFCCs, and traditional classification methods, such as SVMs, for the recognition of environmental sounds, and more specifically sounds of everyday life. Since the application targeted by this work is the remote monitoring of elderly people, we need to use a particular corpus of sounds and consequently a very limited number of classes, given that the tests are in their early stages. As we will see later in the next section, the choice of MFCCs and SVMs methods is not done at random but because the performance they achieve. The second goal is to enrich the reader by a review of the state of the art pertaining to environmental sound recognition and to highlight the major issues and challenges to be resolved.

This paper, aims first at defining the suitable solution for our application according to the findings of the research study, and then, test the proposed solution with a real database and discuss and interpret the results. To accomplish our objective, the paper is structured as follows: we first provide a state-of-the-art overview of sound recognition systems, followed by a synthesis of work on sound recognition systems for distress situation detection purpose in order to see what methods are used and what kind of classes are targeted and to discuss their results, after that we present a survey of studies employing new classification methods based on deep learning for a comparison purpose with the traditional classification methods. Next, we present the general architecture of the sound recognition system. In section 4, we present a brief experimentation and we discuss the results obtained. Finally, the last section concludes the paper with a summary of our findings, conclusions and future research directions.

2. LITERATURE REVIEW

Many research works have been proposed in the field of sound recognition to identify different categories of sounds depending on the intended application. In this section, we present a synthesis of the works and a comparison of the studies carried out in the field of automatic sound recognition. Our comparison is based on the feature extraction methods used as well as the classification methods but we also take into a count the number of classes and consequently the accuracy of the system. Our goal is to show which methods are most commonly used and yield better recognition rates. Classification methods revolve around GMMs (Gaussian Mixture Models), HMMs (Hidden Markov Models), and SVMs. In a previous work [9], we presented a state of the art on environmental sound recognition focusing on the feature extraction methods and classification methods used. In this section, in addition to work on environmental sound recognition, we present work on distress detection and the interest of new classification methods based on deep learning in comparison with classical methods such as GMMs, HMMS and SVMS.

2.1 Overview of research on environmental sound recognition

In the study [10], three classification methods were selected for water sound recognition, namely SVM, KNN (K-Nearest Neighbors) and CNN (convolutional neural network). The features used are extracted from audio fingerprints (20 features). The recognition rate obtained by SVM, which is 98.22%, is higher than the rates obtained by the other two classification models, which are 97.75% and 70.29% for KNN and CNN respectively. The research [11] focused on recognizing environmental sounds to interpret a scene or the context around an audio sensor. The Matching Pursuit (MP) method was selected to extract the most effective frequency-time domain features, and a GMM was used for classification. To illustrate the effectiveness of MP features, tests were conducted using MFCCs, MP features, and a combination of both. The classification accuracy for 14 distinct audio environments was 75.3% with MFCCs, 84.0% with MP features, and 89.7% when combining both. In a previous study by the same group [12], these results were obtained for three different classifiers: 96.6% accuracy for SVM, 94.3% for KNN, and 93.4% for GMM, using forward feature selection. 34 characteristics in all were utilized, including Spectral centroid, spectral bandwidth, the 1st to 12th MFCCs and their standard deviation, etc. An equally significant study [13] proposed a method for location classification using 'audio fingerprints'. This approach utilizes 62 features spanning the temporal, frequency, and statistical domains and tested with two classifiers: Random Forest and SVM. The sound database used for evaluation was sourced from the online collaborative platform Freesound [14], covering 14 different environments. The results indicated classification rates of 84.28% for Random Forest and 91.42% for SVM. A Chi-squared filter is applied for feature selection in a location classification task, reducing thus the number of features from 62 to 15, comprising 11 statistical and 4 frequency features [15]. Using an SVM classifier with 10 classes, the recognition rate exceeded 90%. Muhammad and Alghathbar [16] combined MFCCs, MPEG-7, and zero-crossing rate (ZCR) descriptors to recognize different environments. The use of MPEG-7 descriptors led to better performance compared to using MFCCs alone. The classifier employed was the HMM and experiments indicated that combining MFCCs and MPEG-7 descriptors resulted in superior performance compared to using each feature type individually. Additionally, incorporating ZCR with these descriptors further enhanced performance for certain environment types. By the same group, Muhammad et al. [17] combined MFCCs with 30 MPEG-7 descriptors, the latter are reduced with The FDR (fisher Descriminant Rtaio) method then with the PCA (Principal component analysis) method to get finally 13 MPEG-7 descriptors combined with MFCCs and form an input for a GMM classifier. The results were promising and the combination of MFCCs and MPEG-7 features gives the best results in comparison to the use of each of them separately for certain type of environments. The work [18] dealt with the recognition of impulsive sounds including explosions, glass breaking and screams. The database used for test contains 6 different classes with a total of 822 signals. For the detection algorithm the median filter is used analyzing energy variations and performs well even in noise. GMM and HMM are used for classification, at an SNR of 70 dB the recognition rate was 98%, while at SNR of 0% it was less than 80%. In the study [19], a sound recognition system called AuditHIS is proposed in order to identify sounds produced in the apartment and consequently recognizing performed activities, this system incorporates the system RAPHAEL for speech recognition and identifying distress keywords in the analyzed signal. A GMM classifier with LFCCs features is used for classifying the input signal into speech or sound. The classification step of signals resulted from the GMM classifier is done with a GMM or HMM classifier. good results are achieved by the GMM when the SNR (Signal to Noise Ratio) is less than 10 dB whereas the HMM classifier gibe better results in a noiseless environment. the number of classes is eight. The system achieved an overall performance of 89.76% for accurate sound/speech differentiation and 72.14% for correctly classified sounds. Multiple one-class SVMs were used for the classification of 9 sound classes (gunshots, broken glass, explosions, slamming doors, barking dogs, cries, children's voices and machines and ringing telephones) with a set of audio features namely, ZCR, MFCC, Energy, log energy, SC (Spectral Centroid), SRF (Spectral Roll-off Frequency), and Dynamic Window Composition (DWC) [20]. The correct classification rate is 96.89%. Another trend for environmental sound recognition that used PR descriptors for the classification of sounds into 7 classes namely: screams, broken glass, gunshots, rain, dog barking, restaurant noise and engine sounds [21]. The used classifiers are SVM classifier with linear and Gaussian kernel, a neural network classifier with radial basis function (RBF) and a classifier based on the nearest neighbor method. The results showed that combining PR descriptors and MFCCs gives always the best results for all the types of classifiers and MFCCs features outperforms the PR features. The recognition rates are 88.7% for SVM with Gaussian kernel, 81.78% for RBF neural network and finally 86.4% for NN classifier. The study [22] recognized environmental sounds using MPEG-7 descriptors and temporal Zero Crossing (ZC) with a KNN classifier. Performance improved by increasing the number of training files and decreasing the number of samples per file. You and Li [23] introduced TESPAR (Time Encoded Signal Processing and Recognition) for environmental sound recognition. This method is notable for its low computational requirements compared to other techniques. TESPAR was tested using a database from Freesound, and its performance was compared with an MFCC-based system using an SVM classifier on the same dataset. The results indicated that TESPAR is more effective in noisy environments and has significantly shorter computation times than SVM. Finally, a recent work [24] used SVM classifier for the classification of sound sources in a domestic environment rather than solutions based on deep learning models in order to get high accuracy with low complexity costs. Features such as spectral spread and GTCC (Gammatone Cepstral Coefficients) are extracted and the accuracy is 80% in the validation phase and 60% in a real-time environment.

This section is considered important in this study, given that the most frequently asked question in the field of environmental sound recognition concerns two essential points: feature extraction methods and the classification methods adopted in sound recognition systems. In addition to these last two points, the sound classes targeted by each application also constitute an important criterion in the quality of the system in question because a high recognition rate for a small number of classes cannot be the same for a large number of classes. In this comparative study, we found that most sound recognition systems either rely on MFCCs or combine them with other parameters to increase recognition rates. Time-frequency domain features are also widely used for environmental sound recognition. Previous comparative studies [1, 11], claimed that MFCCs work well for structured sounds such as speech and music [11] and they are also the most widely used in speech and sound recognition applications [1], but their performance degrades in the presence of noise [1, 11]. In addition, MFCCs are not effective for analyzing noise-like signals with a flat spectrum.

Regarding classification methods, we also note that this study focused solely on three types of classifiers: GMM, HMMs and SVMs to compare their recognition rates. Not to mention some studies that tested KNN and ANN. One motivation for this comparison is the choice of the appropriate method for our application and needs which is remote telemonitoring of elderly and disabled.

From this summary we can see that most works use SVM, and it proves to be the most powerful method when compared with other traditional methods. SVM with a Gaussian kernel also shows very satisfactory results when compared with other kernels such as the linear kernel. Not only the use of SVMs in most ESR (Environmental sound recognition) work has encouraged us to use this method in this work, but also its solid theoretical foundation and its capacity for discrimination and generalization are also a cause for choosing the SVM as the first solution for our classification system of everyday life sounds. Lastly, it is also important to point out that in all the works presented above the database used differs from one work to another, as do the classes of sounds to be recognized, which makes performance comparison a difficult task.

2.2 Works on distress situation detection

As for environmental sound recognition systems, we were curious to know the features and the type of classifiers that are mostly used for a specific task which is tele-surveillance of elderly or disabled for distress situation detection via sound recognition. In these systems, some categories of sounds may indicate a distress situation like screams, gunshots, explosions and we often use the term *abnormal sounds* for this type of sounds.

In this section, we review some research studies on remote monitoring systems, specifying the nature of the environment, the number of classes, the types of sounds studied, the classification methods and the recognition rate.

Kuklyte et al. [25] studied the recognition of abnormal events in a noisy environment using MFCCs and HMMs. Four classes were targeted: explosion, gunshots, screams as abnormal or distress sounds and subway noise as a normal event. The correct classification rate was 93.3%. Another remote monitoring work [26] presented a hybrid solution for remote monitoring aimed at detecting crime in an elevator. The proposed system is composed of two subsystems: the first is a supervised classifier and the second is an unsupervised audio analyzer. The aim of the latter is to detect other suspicious sounds not supported by the first subsystem, with the possibility of updating the model by adding new classes of suspicious sounds. This system is based on GMM. The database is composed of 4 classes: clacking, footstep sounds, non-neutral speech and normal speech. The parameters used are 12 MFCCs for an 8-millisecond frame, and the study was carried out on 126 audio clips with suspicious sounds and 4 clips without events. The recognition rate obtained by GMM was 85%. In the study [27], a system for detecting distress situations in a public place was presented. The sound classes are shouting, noise and gunshots, and the system uses two binary GMMs in parallel to distinguish between shouting and noise, and gunshots and noise respectively. Each frame is classified by both classifiers simultaneously. The final decision is made by computing the logical OR. Different types of parameters are used: temporal, spectral, perceptual and correlation. Another work [28] also described an audio monitoring system in a public transport vehicle, which is a noisy environment. Different sounds can occur, and the work is based on the following 5 scenarios or sounds: a fight between two or more men, between two or more women, between men and women, armed robbery, and purse snatching. The experiment was limited to the detection of screams. The acoustic parameters used were MFCCs, LPCs, energy, and PLPCs, and for classification SVM and GMM were used to compare performance and detect screams in this environment. The recognition rate achieved was 75% for the detection of screams, and 98% for the detection of events not containing screams. Experimentation showed that the SVM classifier with the use of PLPs gave the best performance. Valenzise et al. [29] described an audio monitoring system for the detection and localization of abnormal events in a public place, such as shouting and gunshots. The system uses two GMM classifiers working in parallel to discriminate, respectively, shouts from noise and gunshots from noise. The number of features used is 13 for the cries/noise classification and 14 for the gunshots/noise classification after a parameter selection stage. An accuracy of 93% was obtained, with a false rejection rate of 5% when the SNR is 10dB. The work [30] enabled the detection of road accidents by identifying dangerous situations such as tire skidding and car accidents. SVM with a linear kernel was used, and the results showed that the MFCCs and Bark parameters are the best for different SNRs. The average accuracy is 78.95% at a maximum distance of 120 meters. Cakır et al. [31] proposed to apply a CRNN (a combination of CNN and recurrent neural networks (RNN)) to a polyphonic sound event detection (SED) task. The term polyphonic means that the system can handle the existence of several sounds at the same time, as opposed to the monophonic word. The results provided by the system are promising for the four databases tested. Furthermore, the results show a significant improvement with the introduction of deep learning methods. CRNN clearly outperforms previous methods (HMMs, GMMs) and offers considerable improvement compared to other neural network approaches. Min et al. [32] presented a system for detecting indoor emergency events using CNNs. The sounds in question are: sounds indicating emergency events (explosion, gunshot, broken glass and scream) and a single normal sound (sleep). The acoustic parameters calculated were log-scaled mel-spectrograms. The experiment resulted in an F-score of 77.32%.

For a better understanding of this synthesis, we have summarized it in Table 1.

In this section, we've explored some works on distress situation detection which means the recognition of abnormal sounds indicating danger, such as gunshots, whether indoor or outdoor. Recognition of a few classes of sound is a common aspect of these different projects. Various databases have been used to evaluate the proposed systems. The classification methods used are diverse, including GMMs, HMMS, SVMS and deep learning methods. The recognition rates obtained vary from one application to another and from one method to another in the same work. The aim of this section is to see and compare the methods used for the recognition of environmental sounds and those used for the recognition of distress and danger situations such as screams and gunshots. We note that there are no special methods for feature extraction and classification of distress sounds, and that the same methods used in event sound recognition can be used for distress situation detection.

Table 1. Distress situation detection related works

Reference	Purpose of Work (or Environment)	Classes	Methods
[25]	Metro	Explosion, gunshot, screams and subway noise	MFCCs and HMMs
[26]	Elevator	Clacking, footsteps, non-neutral speech and normal speech	MFCCs and GMM
[27]	Public Place	Shouts, noise and gunshots	GMM
[28]	Public Transport Vehicle	Screams	MFCCs, LPCs, energy, and PLPCs, SVM and GMM for classification
[29]	Public Place	Screams, gunshots	GMM
[30]	The Road	Tire skidding and car accidents	MFCCs, bark and SVM
[31]	Indoor and Outdoor Environment (House)	Different DBs and different classes	CRNN-combination of CNN and RNN
[32]	Inside the House	Explosion, gunshot, glass breaking and screaming	CNN

2.3 What about deep learning methods for sound recognition?

More recently, Deep Learning, which refers to artificial neural networks with more than one hidden intermediate layer, has gained popularity and achieved impressive results in various machine learning tasks [33, 34]. It has been highly successful in many fields such as natural language processing, speech recognition, computer vision, image and video analysis, and multimedia. Today, several studies utilize deep learning for automatic sound recognition. In the following, we present some recent works on Sound Event Recognition (SER) that are based on deep learning.

The study [35] addressed emergency events with a critical impact on occupants' health and proposes a deep learning-based sound recognition model to monitor occupants' behaviors and detect potential emergencies. Two classification models were used: Convolutional Deep Neural Network (CNN) and Long Short-Term Memory (LSTM). The developed LSTM in this research is an advanced RNN model designed to overcome the shortcomings of traditional RNNs. Experiments were conducted using audio data collected from real Single-Person Households (SPH) environments and online data sharing websites. Experimental results demonstrated that the developed model could successfully distinguish emergency sound events from regular human activities. The CNN outperformed the LSTM in both emergency sound event classification and occupant behavior monitoring, achieving accuracy rates of 83.9% for CNN and 62.6% for LSTM. Greco et al. [36] proposed a deep learning method named AReN (Audio Event Recognition Network) with 21 layers to automatically recognize events of interest in the context of audio surveillance, such as screams, glass breaks, and gunshots. Input signals are represented by a gammatonegram image; a spectrogram based on a gammatone filter bank. The system was tested on three different databases: SESA, MIVIA audio events, and MIVIA road events, with recognition rates of 91.43%, 99.62%, and 100%, respectively. A comparison in this study between traditional machine learning methodologies and deep learning confirms the effectiveness of the proposed approach. Another work [37] addressed the detection and recognition of continuous audio streams in noisy conditions using deep learning methods. For comparison purposes, the following classifiers were tested on the database: HMM with MFCC, SIF (spectrogram image feature) with SVM, SIF with DNN, SIF with CNN. In the case of isolated sounds, the SIF-CNN system achieved the best performance, followed by SIF-SVM and then SIF-DNN. HMM was the least robust to noise. In the case of a continuous audio stream, SVM performances were highly competitive compared to the CNN system in all cases, even more so than DNN. Nanni et al. [38] proposed a detection system to assist in auscultating heart sounds. Two models were tested in this study, namely CNN and CNN combined with LSTM. The accuracy rates obtained for the two models are 93.07% and 91.06%, respectively. Sigtia et al. [34] proposed a system for detecting baby cries and smoke alarms using deep neural networks (DNNs) and compares the results with Gaussian Mixture Models (GMMs) and Support Vector Machines (SVMs). DNNs offer a higher accuracy rate than SVMs and then GMMs. A more recent work [39] aimed to accomplish two sub-tasks: audio classification into 10 classes and the second involves classifying audio into three categories based on low-complexity solutions. In this work, Delta-DeltaDelta and HPSS parameters were used with four classifier models inspired by VGGNet, ResNet, LCNN, and InceptionNet. Among the four models, the use of Delta-DeltaDelta surpassed the performance of HPSS, and among them, the use of ResNet presented the highest accuracy. In sub-task B, the use of ResNet reduced using the Delta-DeltaDelta function, showed the best performance at 95.38%. Table 2 summarizes all the work presented above.

In this section we have presented an overview of some work on sound recognition using deep learning methods. Although a limited number of works have been presented in this section, we note the importance of this method and its power when compared with traditional methods such as HMM, GMM and SVM. In most works, the recognition rate obtained by these deep learning methods is the highest. As there are different deep learning methods, each method gives a different result, which may be better than the results obtained by other classical methods, or vice versa. As a result, we see the value of introducing these methods in our future tests and work. However, if we think to use our system in edge devices such as smart phones, the traditional methods are more suited than deep learning methods because of the low complexity in calculation and consequently the possibility to design real-time applications as it is mentioned in the study [24].

Table 2. Sound recognition systems based on deep learning methods

Work	Objective	Classification Method	Results Description
[34]	Alarm detection system	-DNNs -GMMs -SVMs	DNNs the best, then SVM, then GMM
[35]	Emergency event detection	CNN LSTM	83.9% for CNN and 62.6% for LSTM
[36]	Audio surveillance	AReN (Audio Event Recognition Network)	91.43%, 99.62% and 100% for three different DB
[37]	Detection and recognition of a continuous Audio stream in noisy conditions	-HMM with MFCC -SIF (spectrogram image feature) with SVM -SIF with DNN -SIF with CNN	-SIF-CNN then SIF-SVM then SIF-SVM -SVM performance is very competitive with CNN, more so than DNN - HMM is the least robust to noise
[38]	Sound detection system to help auscultate Heart sounds	-CNN -CNN combined with LSTM (Long Short-Term Memory)	-93.07% -91.06%
[39]	-Classify audio into 10 classes -Classify audio into three categories	Deltas-DeltaDeltas and HPSS with: VGGNet, ResNet, LCNN, and InceptionNet	ResNet and Deltas-DeltaDeltas presented the best performances at 95.38%

2.4 Concluding remarks

In conclusion, from what has been presented in the previous sections, we see the interest of traditional classification methods (such as SVMs, GMMs, ...) for sound recognition, and SVM almost always shows better results. On the other hand, classification methods based on deep learning also show very satisfactory results, most of the time surpassing those obtained by traditional methods, but SVM also remains a competitor for these classifiers. In addition, SVMs are crucial for environmental sound recognition due to their strong performance in distinguishing between different sound classes, even in high-dimensional feature spaces. They are effective for handling complex, non-linear data often found in environmental sounds data because of their ability to maximize the margin between classes. Moreover, SVMs are robust with smaller datasets, frequently achieving high accuracy without needing extensive data, making them ideal for practical sound recognition applications. We synthetize also, that most sound recognition systems either rely on MFCCs or combine them with other audio features to increase recognition rates. MFCCs are time-frequency domain features and they are particularly valuable in environmental sound recognition as they capture both time and frequency domain information, allowing for a more comprehensive analysis of sound patterns over time and across different frequency ranges. Finally, regarding deep learning methods, they are powerful for environmental sound recognition and generally, outperforms the traditional methods. However, they are computationally intensive, and deploying them in real-time applications requires high processing power. In addition, they require large datasets for training. Therefore, the choice of SVM and MFCCs initially represent the best option considering the criteria and objectives of our application.

3. GENERAL ARCHITECTURE OF A SOUND RECOGNITION SYSTEM

To the best of our knowledge, according to the work done on sound recognition for a remote monitoring system, there are two ways of recognizing sounds:

Either by integrating a categorization phase that distinguishes a speech signal from a sound-type signal, then depending on the type of output sound (either sound or speech) the signal is directed to be processed by the appropriate system [40, 41].

A second solution consists of processing the input signal directly by the classifier and considering speech as a class in addition to the sound classes to be recognized [42]. In this second solution, when the recognized sound class is that of speech, another system can be used to recognize the input speech.

Another solution could be to integrate new classes corresponding to speech, by setting distress keywords such as: help, SOS, etc. However, in this solution, the system's response time would be longer, given the increase in the number of classes to be recognized and, consequently, the increase in the necessary processing time.

From these existing works and architectures of sound recognition systems for remote monitoring applications, we can deduce that the solution adopting a binary classifier to distinguish between the two existing sound categories (speech and sound) is the most realistic. However, it must be approached with caution because its results will influence the entire system.

Similarly, for the sound detection phase, it is a very critical phase because a signal that exists but is not detected by this module can put the entire system at risk [43]. On the other hand, this module will bring time and processing gains by reducing the number of samples to be processed by eliminating signals with periods of silence, and even low-energy signals. Therefore, all these proposals, with their advantages and disadvantages, lead us to the architecture presented in the Figure 1, which is considered the most appropriate solution for us after comparison, and it provides an overview of the sound recognition system.

As we discussed in the previous section, SVMs are one of the most powerful classification methods. Therefore, our first motivation was to explore this technique in combination with MFCCs, which also yield encouraging results in most sound recognition works, including our initial experiments and especially in the present study. However, other methods will be tested in our future work for a comparison purpose. The future tests will be done on the methods that take into consideration the properties of environmental sounds like stationarity, non-stationarity, impulsive behavior, etc. Figure 2 shows the architecture of the sound classification system.

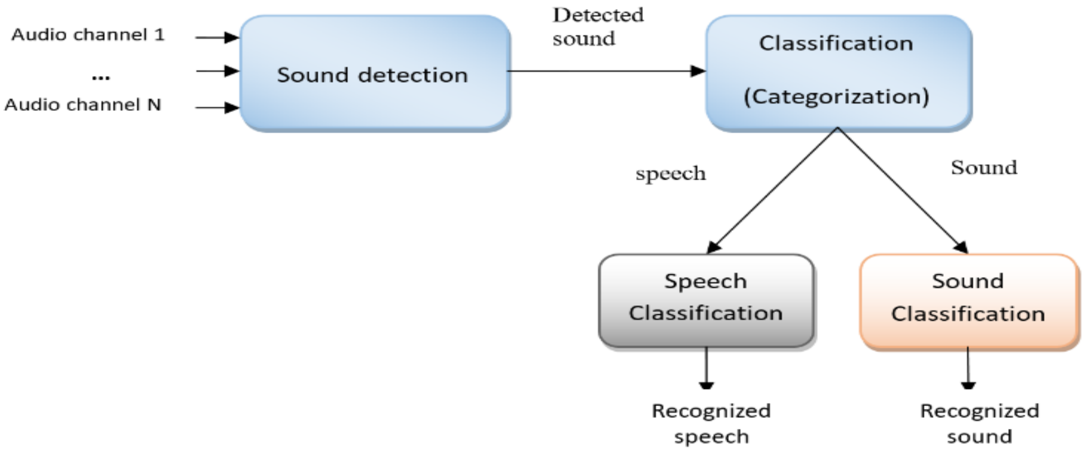

Figure 1. Overview of the sound recognition system

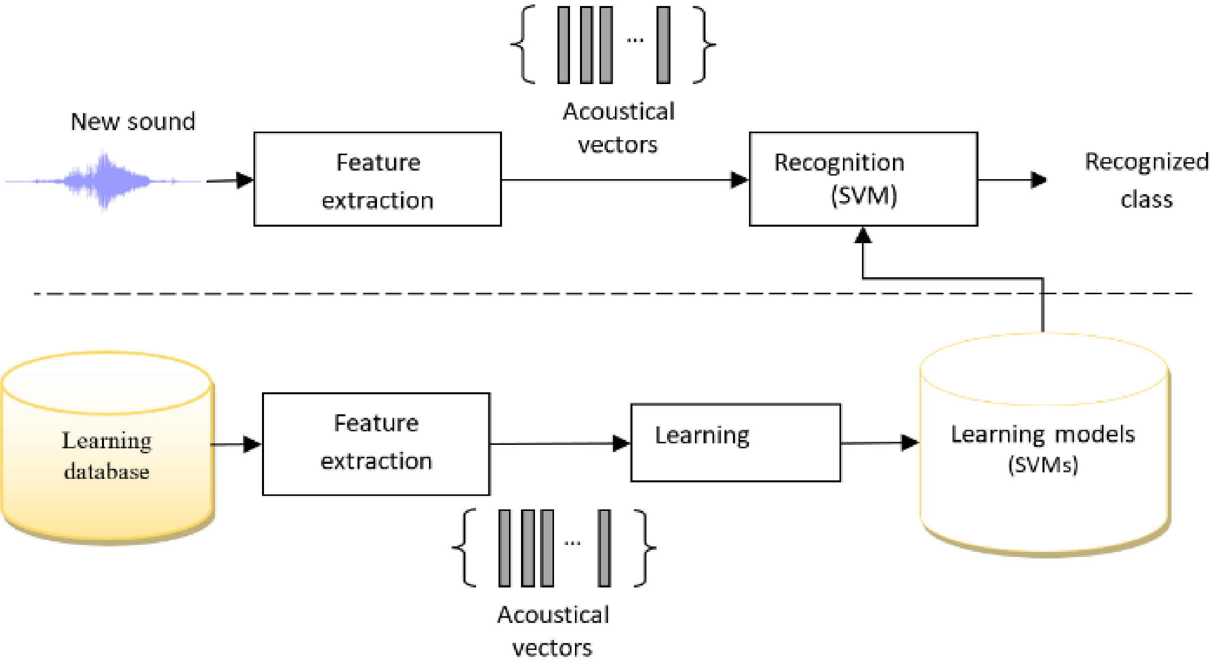

Figure 2. Architecture of the sound classification system showing the main modules

4. EXPERIMENTATION

In this section we present our first experiment for the recognition of a few classes of everyday sounds using MFCCs and an SVM-based classifier.

Indeed, in our previous work [44], we proposed a corpus of everyday life sounds with a development of a small database that was tested using Euclidean distance based classifier and ZCR and energy parameters. In this work, the noise-of-life database [45] was used for the recognition of a few classes of sounds. Indeed, the selection of this database is specifically intended for a future performance comparison between our system and others that use the same database. We have opted for 2 case studies of our sound corpus:

-The first consists in applying our SVM to 3 classes of sound, namely screams, broken glasses and dishwashing sounds.

-The second is to increase the number of classes to be recognized from 3 to 7, in order to compare results and gain a better understanding of the system's behavior for a higher number of classes.

4.1 Corpus of everyday life sounds

The corpus of everyday life sounds provided in the study [44], allows us to understand the state of the inhabitant according to the class of sound detected by the system. These sounds can be Critical sounds or Normal sounds. Critical sounds also called abnormal sounds are the sounds indicating a distress situation of the inhabitant like screaming. Normal sounds can be useful sounds or disturbing sounds; the useful sounds are sounds that can help us to detect the inhabitant activity or to detect a possible distress situation when combined with another information from another sensor. Whereas disturbing sounds, they are sounds that are considered as noise, like electrical devices sounds. Examples for each category of sound are presented in Table 3.

Table 3. Everyday life sounds corpus

Critical Sounds	Normal Sounds	
	Useful Sounds	**Disturbing Sounds**
Screaming Falls of objects Glass breaking A long silence	Sounds of dishes, Doors closing, Doors opening, Door slamming, Sounds of footsteps, Water flow, Coughing, Yawning	TV, Radio, Phone ringing, Electrical devices sounds, External noise

4.2 Dataset

Freesound [14] is a database that has been widely used in different research works on audio recognition, it contains more than 160,000 audio samples. It was used for the benchmarking of diverse studies [11, 13, 15, 21, 23]. The dataset we used here is downloaded from the study [45], which is developed and employed in Sehili’s PhD thesis [46], where a smaller version of this database was collected from web and the second part is extracted from Freesound database.

Indeed, in this work, no data augmentation or class balancing techniques were applied. The dataset was used in its original form without any modifications. Although there may be an imbalance between normal and abnormal sounds; the decision was made to rely on the natural distribution of the data to evaluate the performance of the system without introducing synthetic data or balancing adjustments.

The tested sound classes are the followings: screams, sounds of dishes, glass breaking, door opening, door slamming, coughing and water flow.

The number of samples per sound class is on average around 100 as described in Table 4. The sampling frequency is 16kHz, with a .wav format. We have done a decomposition of the learning base into 20% for testing and the rest for learning.

Table 4. Number of samples per sound class

Sound Class	Number of Samples
dishes	170
screams	193
glass breaking	101
door opening	21
door slamming	114
water flow	54
coughing	62

4.3 Classification with SVM

4.3.1 MFCCs features

The MfCCs are calculated after transforming the signal into the spectral domain, we take the 12 MFCCs as an input for the SVM. Since environmental sounds contain very short, non-stationary and impulsive sounds that change their acoustic characteristics very quickly over time, the window length must be small. We applied a rectangular window of 20-30 milliseconds with 50% overlap for each sound to calculate the acoustic parameters.

4.3.2 SVM classification method

SVM is regarded as one of the most effective approaches for tackling complex classification problems and known as a maximum margin classifier [21, 47] due to its power to find the optimal separating hyperplane that maximizes the distance between the closest points of classes and the separating hyperplane. In addition, an important characteristic of SVMs is that they are not very sensitive to the dimension of the descriptor vectors [20]. More information on SVM principle can be found in the studies [47, 48].

In this experiment we used an SVM classifier with different kernels: linear SVM, polynomial kernel, sigmoid kernel and Radial Basis function (RBF) kernel in order to compare them.

The *C* parameter regulate the trade-off between maximizing the margin and reducing the classification error. The optimal value of the parameter *C* is fixed by using grid search, and the same for degree, and gamma, coef0 and then we applied cross validation technique to obtain the best values that optimize the models.

4.3.3 Main results

In the case of 3 classes namely: screams, sounds of dishes, and glass breaking, the recognition rates achieved using three different kernels, in addition to the linear kernel, are given in Table 5 and presented in the Figure 3.

The Confusion Matrix for the case of SVM with a Gaussian kernel (screams, sounds of dishes and broken glass) is presented in Figure 4.

For the case of 7 classes the sounds to recognize are: glass breaking, door opening, door slamming, shouting, washing up, coughing, water dripping. Table 6 and Figure 5 show the recognition rates for the SVM with the different kernels.

Table 5. SVM classifier recognition rates for four different kernels: 3 sound classes

Kernels Rate (%)	Polynomial	Sigmoid	RBF (Gaussian)	Linear
Recognition rate	83.03%	59.61%	96.28%	100.00%

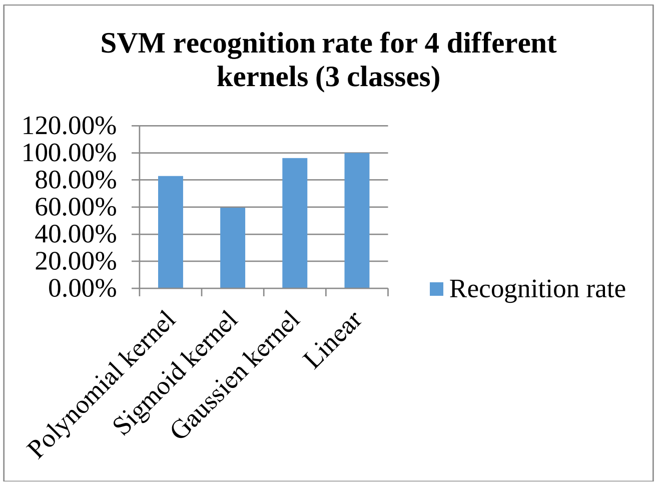

Figure 3. SVM recognition rates for different kernels: 3 classes

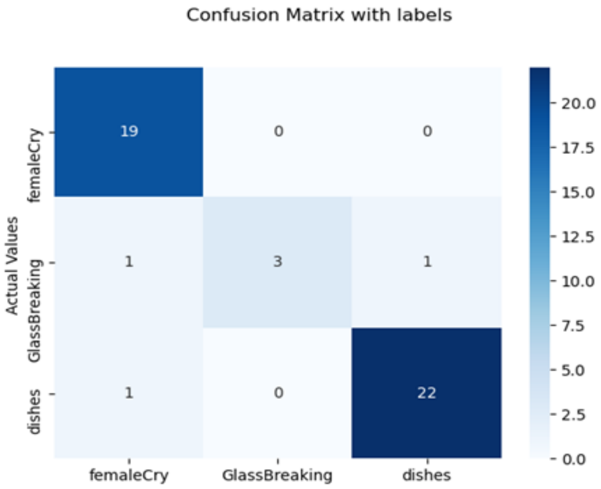

Figure 4. Confusion Matrix for SVM with Gaussian kernel (3 classes)

Table 6. SVM recognition rates for different kernels: 7 sound classes

Kernels Rate (%)	Polynomial	Sigmoid	RBF (Gaussian)	Linear
Recognition rate	72.12%	36.54%	74.04%	95.19%

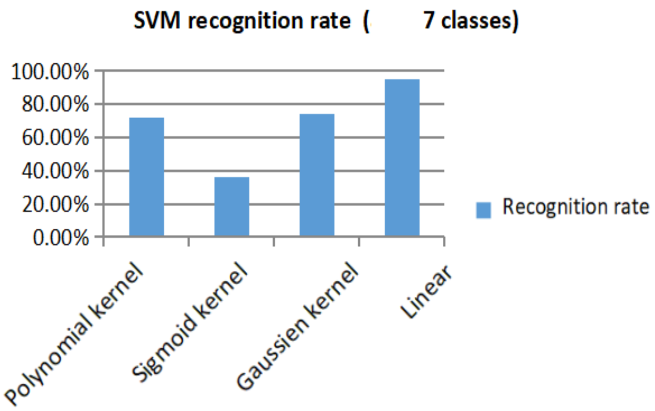

Figure 5. SVM recognition rate (7 classes)

4.3.4 Discussion and interpretation

The results obtained in the case of 3 classes are very encouraging (83.03% for the polynomial kernel, the sigmoid kernel 59.61%, the Gaussian kernel 96.28% and the linear SVM 100%), with the linear SVM showing the best results, followed by the Gaussian kernel. This surprised us, as we expected the Gaussian kernel to be better.

Same results as for the 7-class case, except that recognition rates deteriorate (72.12% for the polynomial kernel, 36.54% for the sigmoid kernel, 74.04% for the Gaussian kernel, 95.19% for the linear kernel), but the ranking order of these kernels is still the same; the linear SVM then comes the Gaussian kernel, the polynomial kernel and finally, the sigmoid kernel.

In addition to these results, we also conclude that it's not just the number of classes that has an impact on the recognition rate, but also the nature of the classes to be recognized. For example, in the case of 7 classes, we don't obtain the same results if we replace one class by another (the "door closing" class by the "door slamming" class, for example). This depends on the similarity or dissimilarity of the new class to the existing classes, so the recognition rate may increase or decrease.

How can we explain the obtained results?

Theoretically, the choice between a linear and non-linear kernel is justified by the nature of the data being processed. In other words, for non-linearly separable data, the use of a non-linear kernel is necessary, while for separable data, a linear kernel is sufficient. In our experimentation, we tested various non-linear kernels in addition to the linear kernel, but the latter showed the best performance. This can be attributed to two factors:

(1) The first factor is the small number of samples (tens or hundreds per class) compared to the feature size. According to previous studies and works, the Gaussian kernel can be used when the number of training samples is very high and the number of features is small. Therefore, the linear kernel yields better results when the number of learning examples is small, as is the case in this experimentation.

(2) The second point is the parameter C (C is a regularization parameter to be set by the user), which plays a significant role in the results. Consequently, the value assigned to C will influence the size of the margin and classification errors. In other words, a large value of C leads to a small margin, and conversely, when the value of C is small, the margin will be large.

Finally, we must mention here that in view of the small number of samples in the sound classes and due to the variation of the number of samples from one class to another, it would be better to use data augmentation and class balancing techniques to get more accurate results if we will use the same database. Data augmentation will be employed to increase the diversity and volume of the dataset by generating additional variations of the existing data, which will help improve the robustness and generalization of the model. In the other hand, class balancing techniques will be applied to address any existing imbalances between the different sound classes, ensuring that the model is trained on a more equitable distribution of classes. These enhancements are expected to lead to better performance and more reliable results.

5. CONCLUSIONS AND PERSPECTIVES

In this paper, we have presented and analyzed the field of sound recognition by comparing various works and focusing on audio parameters, classification methods, and highlighting targeted sound classes. Subsequently, we introduced a general architecture for a sound recognition system, followed by the sound classification module based on Support Vector Machines (SVMs), which, according to the conducted study, in some cases, yields results comparable to those of new deep learning methods.

In the experimentation, MFCCs parameters are used as input for the SVM, and we compared the results for different kernels, including the Gaussian kernel, polynomial kernel, sigmoid kernel, and linear SVM. In this work, we initially targeted distress sounds, such as glass breaking and screams. We then, extended the tests to cover other types of sounds to observe the system's behavior and discuss the results.

The results obtained for 3 classes are very promising (83.03% for the polynomial kernel, 59.61% for the sigmoid kernel, 96.28% for the Gaussian kernel, and 100% for the linear SVM). In the case of 7 classes, recognition rates decline (72.12% for the polynomial kernel, 36.54% for the sigmoid kernel, 74.04% for the Gaussian kernel, and 95.19% for the linear SVM). However, the ranking order of these kernels based on the rate of correct classification remains the same: linear SVM, followed by the Gaussian kernel, polynomial kernel, and finally, the sigmoid kernel.

This study offers several practical implications for real-world applications. First, the literature review provides a comprehensive overview of existing methodologies and advancements in sound recognition technology and thus, supports further research and development and suggests areas for improvements. The sound recognition system can be integrated into various monitoring applications, such as security systems, healthcare environments and industrial settings. This depends on the sounds to recognize; distress sounds for health care, alarms in industries, etc. Moreover, the integration of the system in smart environments such as smart homes and cities by facilitating for example, alert systems based on the sounds detected.

Despite the promising results of this study, there are still several challenges to overcome. The first challenge is the lack of a standard database for benchmarking the variety of research studies; validating systems with the same database provides the opportunity to compare the proposed solutions and thus, identifying the most effective methods. Secondly, given the variety of sounds in the environment and their diverse natures; stationary, non-stationary and impulsive, the choice of audio features and classification methods that can take into account all these differences is a challenging and critical task. This leads us to view the sound signal as a structure rather than just a sound, moving away from its representation as a frame, as mentioned in the study [49], and we must therefore test other methods like end-to-end models carefully, because in spite of their improvement of accuracy they might acquire more computational resources. Moreover, resorting to new deep learning methods requires working with larger databases, and this can be achieved by data augmentation for small databases and especially those containing sounds that are difficult to acquire or to reproduce like the fall of persons and explosions. Also, applying balancing techniques will address any existing imbalances between the different sound classes, ensuring that the model is trained on a more equitable distribution of classes and thus, improving accuracy.

Finally, real-time applications like audio surveillance and security systems require fast, real-time recognition. Achieving accurate results in real time, especially with limited computational resources, is challenging. This often requires efficient algorithms that balance accuracy with speed, but this trade-off can limit recognition quality. For this reason, the choice of the classification method and the audio features is a critical task and this is why we have undertaken this research.

In conclusion, our future work should focus on:

- Enhancing the system's ability to accurately recognize sounds in noisy or complex environments.
- Improvement of accuracy and precision by exploring advanced machine learning techniques.
- Optimizing the solution for faster and real-time processing and scaling the system to manage high-throughput data streams.
- Addressing class imbalance and using data augmentation techniques to increase the diversity of sound samples in the dataset or using a more standard database.
- Integrating the sound recognition system with other technologies, such as IoT devices or AI-driven applications.

REFERENCES

[1] Sharan, R.V., Moir, T.J. (2016). An overview of applications and advancements in automatic sound recognition. Neurocomputing, 200: 22-34. https://doi.org/10.1016/j.neucom.2016.03.020

[2] Zhang, L., Shi, Z., Han, J. (2020). Pyramidal temporal pooling with discriminative mapping for audio classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28: 770-784. https://doi.org/10.1109/TASLP.2020.2966868

[3] Jung, J.W., Heo, H.S., Shim, H.J., Yu, H.J. (2020). Knowledge distillation in acoustic scene classification. IEEE Access, 8: 166870-166879. https://doi.org/10.1109/ACCESS.2020.3021711

[4] Zhang, T., Wu, J. (2019). Constrained learned feature extraction for acoustic scene classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(8): 1216-1228. https://doi.org/10.1109/TASLP.2019.2913091

[5] Istrate, D., Vacher, M., Castelli, E., Nguyen, C.P. (2004). Sound processing for health smart home. In Proceedings of the International Conference on Smart Homes and Health Informatics, pp. 41-48.

[6] Bregman, A.S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press.

[7] Abayomi-Alli, O.O., Damaševičius, R., Qazi, A., Adedoyin-Olowe, M., Misra, S. (2022). Data augmentation and deep learning methods in sound classification: A systematic review. Electronics, 11(22): 3795. https://doi.org/10.3390/electronics11223795

[8] Chachada, S., Kuo, C.C.J. (2013). Environmental sound recognition: A survey. 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013.

[9] Abdoune, L., Fezari, M. (2017). Feature extraction for everyday life sounds. In 5th International Conference on Control & Signal Processing(CSP-2017) Proceeding of Engineering and Technology-PET, 26: 186-191.

[10] Hang, T., Feng, J., Li, X., Yan, L. (2019). Water sound recognition based on support vector machine. In Proceedings of the 13th International Conference on Ubiquitous Information Management and Communication (IMCOM). Springer International Publishing. Springer, Cham, 13: 986-995. https://doi.org/10.1007/978-3-030-19063-7_77

[11] Chu, S., Narayanan, S., Kuo, C.C.J. (2009). Environmental sound recognition with time-frequency audio features. IEEE Transactions on Audio, Speech, and Language Processing, 17(6): 1142-1158. https://doi.org/10.1109/TASL.2009.2017438

[12] Chu, S., Narayanan, S., Kuo, C.C.J., Mataric, M.J. (2006). Where am I? Scene recognition for mobile robots using audio features. In 2006 IEEE International Conference on Multimedia and Expo, Toronto, ON, Canada, pp. 885-888. https://doi.org/10.1109/ICME.2006.262661

[13] Delgado-Contreras, J.R., García-Vázquez, J.P., Brena, R.F. (2014). Classification of environmental audio signals using statistical time and frequency features. In 2014 International Conference on Electronics, Communications and Computers (CONIELECOMP), Cholula, Mexico, pp. 212-216. https://doi.org/10.1109/CONIELECOMP.2014.6808593

[14] Freesound. Find any sound you like, https://freesound.org/, accessed on 2023.

[15] Delgado-Contreras, J.R., Garćıa-Vázquez, J.P., Brena, R.F., Galván-Tejada, C.E., Galván-Tejada, J.I. (2014). Feature selection for place classification through environmental sounds. Procedia Computer Science, 37: 40-47. https://doi.org/10.1016/j.procs.2014.08.010

[16] Muhammad, G., Alghathbar, K. (2009). Environment recognition from audio using MPEG-7 features. In 2009 Fourth International Conference on Embedded and Multimedia Computing, Jeju, Korea, pp. 1-6. https://doi.org/10.1109/EM-COM.2009.5402978

[17] Muhammad, G., Alotaibi, Y.A., Alsulaiman, M., Huda, M.N. (2010). Environment recognition using selected MPEG-7 audio features and mel-frequency cepstral coefficients. In 2010 Fifth International Conference on Digital Telecommunications, Athens, Greece, pp. 11-16. https://doi.org/10.1109/ICDT.2010.10

[18] Dufaux, A., Besacier, L., Ansorge, M., Pellandini, F. (2000). Automatic sound detection and recognition for noisy environment. In 2000 10th European Signal Processing Conference, Tampere, Finland, pp. 1-4.

[19] Vacher, M., Fleury, A., Portet, F., Serignat, J.F., Noury, N. (2010). Complete sound and speech recognition system for health smart homes: Application to the recognition of activities of daily living. New Developments in Biomedical Engineering, pp. 645.

[20] Rabaoui, A., Davy, M., Rossignol, S., Lachiri, Z., Ellouze, N., équipe SequeL, I.F. (2007). Sélection de descripteurs audio pour la classification des sons environnementaux avec des SVMs mono-classe. In Actes du 21eme Colloque GRETSI: Traitement du Signal et des Images (GRETSI'07).

[21] Uzkent, B., Barkana, B.D., Cevikalp, H. (2012). Non-speech environmental sound classification using SVMs with a new set of features. International Journal of Innovative Computing, Information and Control, 8(5): 3511-3524.

[22] AlQahtani, M.O., Muhammad, G., Alotaibi, Y.A. (2010). Environment sound recognition using zero crossing features and MPEG-7. In 2010 Fifth International

Conference on Digital Information Management (ICDIM), Thunder Bay, ON, Canada, pp. 502-506. https://doi.org/10.1109/ICDIM.2010.5664645

[23] You, G., Li, Y. (2012). Environmental sounds recognition using tespar. In 2012 5th International Congress on Image and Signal Processing, Chongqing, China, pp. 1796-1800. https://doi.org/10.1109/CISP.2012.6469781

[24] Jesudhas, P.P., Ranjan, P.V. (2024). A novel approach to build a low complexity smart sound recognition system for domestic environment. Applied Acoustics, 221: 110028. https://doi.org/10.1016/j.apacoust.2024.110028

[25] Kuklyte, J., Kelly, P., Ó Conaire, C., O'Connor, N.E., Xu, L.Q. (2009). Anti-social behavior detection in audio-visual surveillance systems. In the Workshop on Pattern Recognition and Artificial Intelligence for Human Behavior Analysis, Reggio Emilia, Italy, 2009.

[26] Radhakrishnan, R., Divakaran, A., Smaragdis, A. (2005). Audio analysis for surveillance applications. In IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005., New Paltz, NY, USA, pp. 158-161. https://doi.org/10.1109/ASPAA.2005.1540194

[27] Ntalampiras, S., Potamitis, I., Fakotakis, N. (2009). On acoustic surveillance of hazardous situations. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. Taipei, Taiwan, pp. 165-168. https://doi.org/10.1109/ICASSP.2009.4959546

[28] Rouas, J.L., Louradour, J., Ambellouis, S. (2006). Audio events detection in public transport vehicle. In 2006 IEEE Intelligent Transportation Systems Conference, Toronto, ON, Canada, pp. 733-738. https://doi.org/10.1109/ITSC.2006.1706829

[29] Valenzise, G., Gerosa, L., Tagliasacchi, M., Antonacci, F., Sarti, A. (2007). Scream and gunshot detection and localization for audio-surveillance systems. In 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, London, UK, pp. 21-26. https://doi.org/10.1109/AVSS.2007.4425280

[30] Foggia, P., Petkov, N., Saggese, A., Strisciuglio, N., Vento, M. (2016). Audio surveillance of roads: A system for detecting anomalous sounds. IEEE Transactions on Intelligent Transportation Systems, 17(1): 279-288. https://doi.org/10.1109/TITS.2015.2470216

[31] Cakır, E., Parascandolo, G., Heittola, T., Huttunen, H., Virtanen, T. (2017). Convolutional recurrent neural networks for polyphonic sound event detection. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(6): 1291-1303. https://doi.org/10.1109/TASLP.2017.2690575

[32] Min, K., Jung, M., Kim, J., Chi, S. (2018). Sound event recognition-based classification model for automated emergency detection in indoor environment. In Advances in Informatics and Computing in Civil and Construction Engineering: Proceedings of the 35th CIB W78 2018 Conference: IT in Design, Construction, and Management. Cham: Springer International Publishing. Springer, Cham, pp. 529-535. https://doi.org/10.1007/978-3-030-00220-6_63

[33] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553): 436-444. https://doi.org/10.1038/nature14539

[34] Sigtia, S., Stark, A.M., Krstulović, S., Plumbley, M.D. (2016). Automatic environmental sound recognition: Performance versus computational cost. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(11): 2096-2107. https://doi.org/10.1109/TASLP.2016.2592698

[35] Kim, J., Min, K., Jung, M., Chi, S. (2020). Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition. Building and Environment, 181: 107092. https://doi.org/10.1016/j.buildenv.2020.107092

[36] Greco, A., Petkov, N., Saggese, A., Vento, M. (2020). AReN: A deep learning approach for sound event recognition using a brain inspired representation. IEEE Transactions on Information Forensics and Security, 15: 3610-3624. https://doi.org/10.1109/TIFS.2020.2994740

[37] McLoughlin, I., Zhang, H., Xie, Z., Song, Y., Xiao, W., Phan, H. (2017). Continuous robust sound event classification using time-frequency features and deep learning. PloS One, 12(9): e0182309. https://doi.org/10.1371/journal.pone.0182309

[38] Nanni, L., Costa, Y.M.G., Lucio, D.R., Silla, C.N., Brahnam, S. (2017). Combining visual and acoustic features for audio classification tasks. Pattern Recognition Letters, 88: 49-56. https://doi.org/10.1016/j.patrec.2017.01.013

[39] Lee, Y., Lim, S., Kwak, I.Y. (2021). CNN-based acoustic scene classification system. Electronics, 10(4): 371. https://doi.org/10.3390/electronics10040371

[40] Vacher, M., Istrate, D., Besacier, L., Serignat, J.F., Castelli, E. (2023). Life sounds extraction and classification in noisy environment. In 5th IASTED-SIP, Honolulu, Hawaii, USA.

[41] Istrate, D., Vacher, M., Serignat, J.F. (2008). Embedded implementation of distress situation identification through sound analysis. The Journal on Information Technology in Healthcare, 6(3): 204-211.

[42] Wang, J.C., Lee, H.P., Wang, J.F., Lin, C.B. (2008). Robust environmental sound recognition for home automation. IEEE Transactions on Automation Science and Engineering, 5(1): 25-31. https://doi.org/10.1109/TASE.2007.911680

[43] Istrate, D. (2003). Détection et reconnaissance des sons pour la surveillance médicale. Doctoral Dissertation, Institut National Polytechnique de Grenoble-INPG.

[44] Abdoune, L., Fezari, M. (2016). Everyday life sounds database: Telemonitoring of elderly or disabled. Journal of Intelligent Systems, 25(1): 71-84. https://doi.org/10.1515/jisys-2014-0110

[45] Sehili, A. (2017). Noise-of-life. https://github.com/amsehili/noise-of-life.

[46] Sehili, M.E.A. (2013). Reconnaissance des sons de l'environnement dans un contexte domotique. Doctoral Dissertation, Institut National des Télécommunications.

[47] Burges, C.J. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2): 121-167. https://doi.org/10.1023/A:1009715923555

[48] Cortes, C. (1995). Support-Vector networks. Machine Learning, 20: 273-297.

[49] Krstulović, S. (2018). Audio event recognition in the smart home. Computational Analysis of Sound Scenes and Events. Springer, Cham, 335-371. https://doi.org/10.1007/978-3-319-63450-0_12